

Hanna Celach

Politechnika Lubelska, Lublin

Trzech uczonych, trzy bazy, jeden bibliotekarz. O pożytkach i metodzie uczelnianej analizy cytowań

Three scholars, three databases, one librarian. On the uses and method of citation analysis for a university. For many years Science Citations Index, published by the Institute of Scientific Information in Philadelphia served for analysis of scientific output by the way of citations count and tracking. In Lublin University of Technology we started doing so in 1995. It was the only citations database available at that time, and although this school's authors were hardly present in it, and the searching functionalities were far from perfect, we continued the screening ever since. In the year 2004 SCOPUS database and Google Scholar search engine showed up, both capable to compete with SCI. This paper is an attempt to compare these three tools. Three authors were selected for the test: two chemists (electrochemistry and environment engineering) and one physicist. The selection was based mostly on the name (unrepeated within the discipline and without Polish diacritics) and the number of publications – between 50 and 100, so that the material for comparison is not too large but large enough to produce a different picture for each author. All three authors have been contributing to scientific journals for more than 15 years, and the analysis covers the period since 1996 till now.

W ostatnich latach powstaje coraz więcej baz danych, a istniejące wzbogacane są o nowe narzędzia do wyszukiwania informacji i przeprowadzania rozmaitych analiz piśmiennictwa naukowego. Pojawiają się też nowe metody oceny publikacji, nowe współczynniki. Zdania na temat ich celowości czy przydatności są w środowisku naukowym podzielone – ale niezależnie od tego, jak ta dyskusja przebiega, bibliotekarze powinni je znać i umieć wykorzystać w swojej pracy.

O nowych metodach naukometrycznych i bazach cytowań pisze się dużo, ale niewiele jest prac porównawczych – znalazłam tylko jedną publikację polską, porównującą trzy bazy (SCI, Scopus, Scholar Google) pod kątem cytowania w nich prac z jednego tytułu czasopisma [5, s. 839–843]. Znacznie więcej jest publikacji zagranicznych, porównujących zawartość również w innych aspektach, np. występowania opracowań z danej dziedziny wiedzy czy zawartości tytułów. Przedmiotem porównania, w niniejszych badaniach będą następujące bazy:

- Science Citation Index Expanded (SCI-Ex) – produkt firmy Thomson Scientific (dawniej ISI), jako rozszerzenie bazy SCI – najstarszej z baz cytujących i najszerzej dotychczas wykorzystywanej;
- Scopus – nowy produkt firmy Elsevier, na rynku od 2002 r., w Polsce od 2004 r.;
- Scholar Google – produkt firmy Google, testowa wersja wyszukiwarki zasobów internetowych, przeznaczona dla pracowników naukowych i studentów; w listopadzie 2004 r. uruchomiono wersję Beta, która funkcjonuje dotychczas

Tabela 1. Porównanie baz SCI, Scopus i Google Scholar

Zawartość	SCI-Ex	Scopus	Scholar Google
	Aktualnie zawiera ponad 20 mln rekordów Indeksuje 6687 tytułów czasopism z ponad 160 dyscyplin naukowych (m.in. Agriculture Neuroscience Astronomy Oncology Biochemistry Pediatrics Biology Pharmacology Biotechnology Physics Chemistry Plant Sciences Computer Science Psychiatry Materials Science Surgery Mathematics Veterinary Science Medicine Zoology)	Zawiera 29 mln abstraktów z ponad 15 tys.* recenzjowanych tytułów (w tym ok. 12,850 czasopism, 700 materiałów konferencyjnych, 600 wydawnictw branżowych, 500 otwartych czasopism i 125 serii książkowych) ponad 4 tys. wydawców, 265 mln cytowań, 265 mln stron www (Scirus) *dane wydawcy	Nieznana liczba rekordów, źródeł, wydawców Ponad 30 różnych typów dokumentów
Zakres	SCI: 1945– Dostęp mamy od 1996 r.	1996–	Nieznany
Zakres tematyczny	164 dyscypliny naukowe	Multidyscyplinarna	Multidyscyplinarna
Opcje wyszukiwania cytowań	Cytowany autor Cytowana praca (wymaga użycia skrótu czasopisma book, or conference title in which the work appeared) Rok cytowania	Opcja “Basic Search” pozwala na wyszukiwanie za pomocą słów kluczowych lub frazy, wpisanych w pole “References” Opcja “Advanced Search” umożliwia wyszukiwanie według: Cytowanego autora (REFAUTH) Tytułu pracy (REFTITLE) Tytułu źródła (REFSRCTITLE) roku (REFPUBYEAR) strony (REFPAGE) Pole (REF), jest kombinacją pól REFAUTH, REFTITLE, REFSRCTITLE, REFPUBYEAR i REFPAGE.	Z poziomu wyszukiwarki, za pomocą dowolnych słów kluczowych lub nazwiska autora, można wyszukiwanie zawęzić do: “Author” “Publication” “Date” oraz “Subject Areas” Znacznie większe możliwości daje program Publish or Perish [4] – jest to aplikacja zewnętrzna, konieczne jest zainstalowanie programu lokalnie (dostępna na stronie www.harzing.com)
Dostępne narzędzia analizy	do Nie można skorzystać z dostępnych, wbudowanych narzędzi, aby szeregować wyniki wyszukiwania tylko listy publikacji z ilością cytowań	Szeregowanie według autora, roku publikacji, źródła, tematu i typu dokumentu. Analiza cytowań w poszczególnych latach, automatyczne wyliczanie współczynnika h, sporządzanie wykresów (przez Citation Tracker)	Publish or Perish: Szeregowanie wg autora, roku publikacji, źródła, tematu i typu dokumentu. Dokładna statystyka publikacji i cytowań z wyliczeniem wszystkich popularnych wskaźników (h, g, A)

Opcje ściągania i eksportowania danych do innych programów bibliograficznych (np. EndNote and RefWorks)	Tak, ale publikacji cytujących, nie cytowanych – cytowane można przekopiować ręcznie do dokumentu tekstowego i zaimportować je do Excela	Tak	Tak – korzystając z PoP
Dostęp	płatny	Płatny	bezpłatny

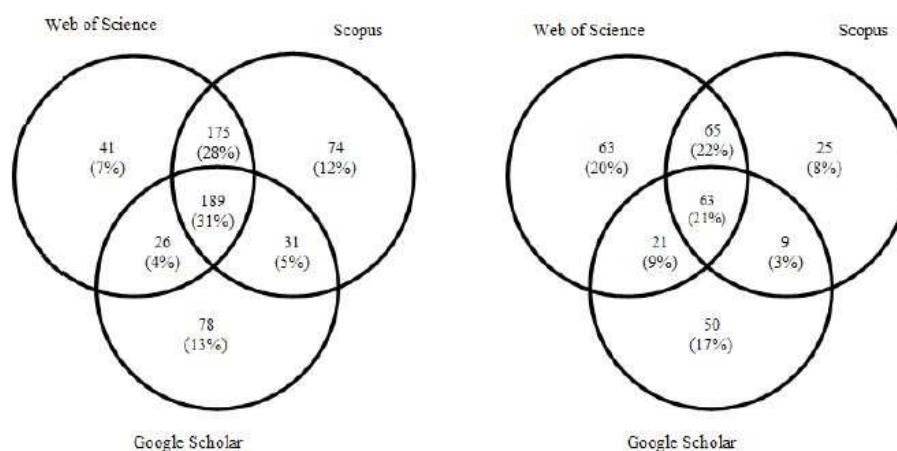
Źródło: Opracowanie własne.

Porównanie przedstawione w tabeli 1, niewiele nam jednak powie, dopóki nie wykonamy konkretnych wyszukiwań i nie porównamy wyników. W każdej z baz wygląda to inaczej, wyniki wyszukiwania oraz sposób ich prezentacji są nieco odmienne. Dopiero w trakcie pracy możemy ocenić zarówno sam sposób porozumiewania się z bazą – zadawania pytań i przeglądanie odpowiedzi, jak i jej zawartość, czyli liczbę uzyskanych wyników i ich kompletność.

Literatura zagraniczna porównująca bazy jest dosyć obszerna, ale koncentruje się głównie na porównaniu baz Scopus i Web of Science (niniejsza próba dotyczy tylko SCI-Ex), nieliczne źródła [1, 2, 3] podają również wyniki ze Scholar Google.

N. Bakkalbasi, K. Bauer, J. Gloger i L. Wang porównali występowanie w trzech bazach cytowań publikacji z dwóch dziedzin [1]. Trochę inaczej wyniki te wyglądałyby, gdyby porównanie zawęzić tylko do bazy SCI.

Figure 1 - The distribution of the unique and overlapped citing articles as returned by the algorithm
Oncology (total=614) **CM Physics (total=296)**



Rys. 1. Porównanie baz według N. Bakkalbasi i in. [1]

Liczba publikacji wspólnych we wszystkich bazach oscyluje w granicach 21%–31% dla tych dziedzin wiedzy. Bardziej zróżnicowana jest liczba publikacji niepowtarzających się – od 7% do 20%, skrajne wartości dotyczą jednej bazy, co może sugerować, że ta dziedzina wiedzy jest po prostu słabo reprezentowana w bazie. Takie badanie bazy ma znaczenie przy planowaniu zakupów, jeśli musimy wybierać jedną z baz, a biblioteka ma określony profil.

Natomiast porównanie baz pod kątem występowania w nich publikacji konkretnego autora ma na celu tylko i wyłącznie wykonanie pełnej analizy cytowań. Wynik tej analizy również we wszystkich bazach jest inny – zależy od tego, co i gdzie dany autor publikuje: jeśli w czasopiśmie z IF i książkach, to najpełniejsza analiza będzie w bazie SCI (zob. tabela 2, autor MK). Widać wyraźnie, że liczba publikacji cytowanych i liczba cytowań różnią się bardzo. Wynika to ze sposobu przeprowadzenia analizy oraz z możliwości wyszukiwania, jakie oferuje nam każda z baz. Wady i zalety baz to zagadnienie osobne, skupię się tu tylko na tych aspektach, które mają bezpośredni wpływ na wynik wyszukiwania cytowań konkretnych autorów.

Wspólną niedoskonałością wszystkich baz jest, oczywiście, uwzględnianie różnej pisowni nazwisk autorów, zwłaszcza nazwisk z polskimi znakami. Baza Scopus wydaje się w tym względzie najbliższa ideałowi – można tu zadawać pytanie używając polskich liter, a otrzymamy wykaz nazwisk pisanych w różnych wersjach. Można wybrać z tej listy nazwisk tylko te, co do których jesteśmy pewni, gdyż obok podane są wszystkie dane pozwalające dokładnie zidentyfikować autora, czyli miejsce zatrudnienia, obszar jego działalności naukowej, a nawet kraj i miasto – nie ma tych opcji w pozostałych dwóch bazach.

Ponieważ Scopus i SCI nie rejestrują książek (z wyjątkiem wydawnictw seryjnych), w przypadku obu baz nie wystarczy skorzystać z najprostszych narzędzi, aby zrobić kompletną analizę cytowań danego autora. W bazie Scopus oprócz „citation tracker” trzeba dodatkowo wykorzystać możliwość wyszukania cytowań w opcji zaawansowanej poprzez pole cytowanego autora (REFAUTH) oraz cytowanego źródła (REFSRCTITLE) i wyniki te dołączyć do wyników cytowań z „citation tracker”.

Niektóre publikacje, zdarza się, są wykazywane kilkakrotnie, mimo że są to te same edycje. Być może sprawiają to niewielkie różnice w opisie bibliograficznym. Ilustruje to następujący przykład jednej z pozycji książkowych prof. Kosmulskiego: wpisując w pole REFSRCTITLE pełny tytuł jego książki *Chemical Properties of Material Surfaces*, otrzymujemy tylko 67 wyników. Widocznie do przeszukania należy inaczej sformułować zapytanie, aby uzyskać pełny wynik, gdyż przy tej pozycji w bibliografii załącznikowej mamy informację, że książka była cytowana 76 razy – klikając ten link, otrzymujemy pełniejszy wynik cytowań:

37.	<input type="checkbox"/> REFSRCTITLE("Chemical Properties of Material* Surface*")	0 Scopus	Edit Save Set Alert
36.	<input type="checkbox"/> REFSRCTITLE("Chemical Properties of Material Surfaces") OR ("chemical properties of materials surfaces") OR ("chemical properties of materials surface")	77 Scopus	Edit Save Set Alert
35.	<input type="checkbox"/> REFSRCTITLE("Chemical Properties of Material Surfaces")	67 Scopus	Edit Save Set Alert

Pozycja z oznaczoną liczbą cytowań:

2. ☐ **Kosmulski, M.**
(2001) *Chemical Properties of Materials Surfaces*. [Cited 76 times](#).
Dekker, New York

Po przejrzeniu wyników okazuje się, że ta sama pozycja jest cytowana tak:

23. ☐ **Kosmulski, M.J.**
Surface charging in absence of strongly adsorbing species
(2001) *Chemical Properties of Material Surfaces*, *Surfactant Science Series*, 102, pp. 65-309.
Edited by A. T. Hubbard. New York: Marcel Dekker

lub tak

38. ☐ **Kosmulski, M.**
(2001) *Chemical Properties of Material Surfaces*, p. 577.
Marcel Dekker Inc, New York, Chapter 5, p

Są to dodatkowe cytowania tej książki, nie uwzględnione w tych 76. Takich cytowań jest sporo i trudno je wyszukać. Drobną różnicą wystąpiła w tytule *Chemical Properties of Materials Surfaces* czy *Material Surface*. Wpisując zapytanie jako frazę, nie możemy użyć prefiksów. Po usunięciu cudzysłowu i użyciu prefiksów: Chemical Properties of Material* Surface* oraz dodaniu dodatkowego pola „Autor” otrzymujemy 85 wyników, wśród których są zapewne również te „niepodwiązane” do wykazanych 76.

Są to tylko subtelności, niemające może dużego znaczenia, ale niekiedy mogą wpłynąć na wielkość współczynnika h – zwłaszcza jeśli podniosą liczbę cytowań publikacji brzegowych.

Drugim ważnym aspektem mającym wpływ na sposób wyszukiwania jest zakres chronologiczny baz. Stosując bazę Scopus do oceny pracowników, których okres aktywności naukowej przypadł również na lata poprzedzające rok 1996, powinniśmy zawsze sprawdzić, czy publikacje tych autorów są cytowane, pomimo że nie wykazuje ich wyszukiwanie z poziomu „Basic Search” według autora. Ilustruje to następujący przykład praca prof. Gładyszewskiego:

5. ☐ Badawi, K.F., Goudeau, P., Pacaud, J., Jaouen, C., Delafond, J., Naudon, A., **Gładyszewski, G.**
X-ray diffraction study of residual stresses modification in cu/w superlattices irradiated by light and heavy ions
 (1993) *Nuclear Instruments and Methods in Physics Research*, 81 B80, pp. 404-407. Cited 18 times.
[View at Publisher](#)

Praca jest cytowana 18 razy, nie widać tego wyniku w analizie wykonanej poprzez „citation tracker”. Oczywiście, ma to wpływ na wartość współczynnika h.

Natomiast w bazie SCI, należy szukać tej publikacji, jako że jest ona z 1993 r., pod pierwszym autorem. W tym przypadku jedna pozycja ma dwie wersje zapisu:

<input type="checkbox"/>	1	BADAWI KF	NUCL I METH PHYS R B	404	1993
<input type="checkbox"/>	16	BADAWI KF	NUCL INSTRUM METH 1	80 404	1993

Analiza cytowań dla trzech autorów, którzy publikują przeważnie w języku angielskim i czasopismach z IF, wykonana bez zagłębiania się w powyższe subtelności, wygląda następująco:

Tabela 2. Trzech autorów w trzech bazach – wyszukiwanie proste*

Baza	MK			DJ			GG		
	liczb publ.	Liczba cytowań	h	Liczba publ.	Liczba cytowań	h	Liczba publ.	Liczba cytowań	h
SCI	129	1168	19	40	727	15	34	191	8
Scopus	93	667	17	30	790	16	28	117	7
Scholar Google	70	458	11	23	334	10	16	49	4

*wyszukiwaniem prostym w tym przypadku nazwano wyszukiwanie bardzo ogólne – z autocytowaniami (w przypadku wszystkich baz), bez publikacji sprzed 1996 r. i bez książek (w przypadku bazy Scopus) oraz bez publikacji, gdzie „mój” autor nie jest pierwszym autorem (dotyczy to publikacji cytowanych w SCI, tych sprzed 1996 r.)

Źródło: Opracowanie własne.

Zupełnie inną analizę otrzymujemy po wprowadzeniu dodatkowych opcji wyszukiwawczych.

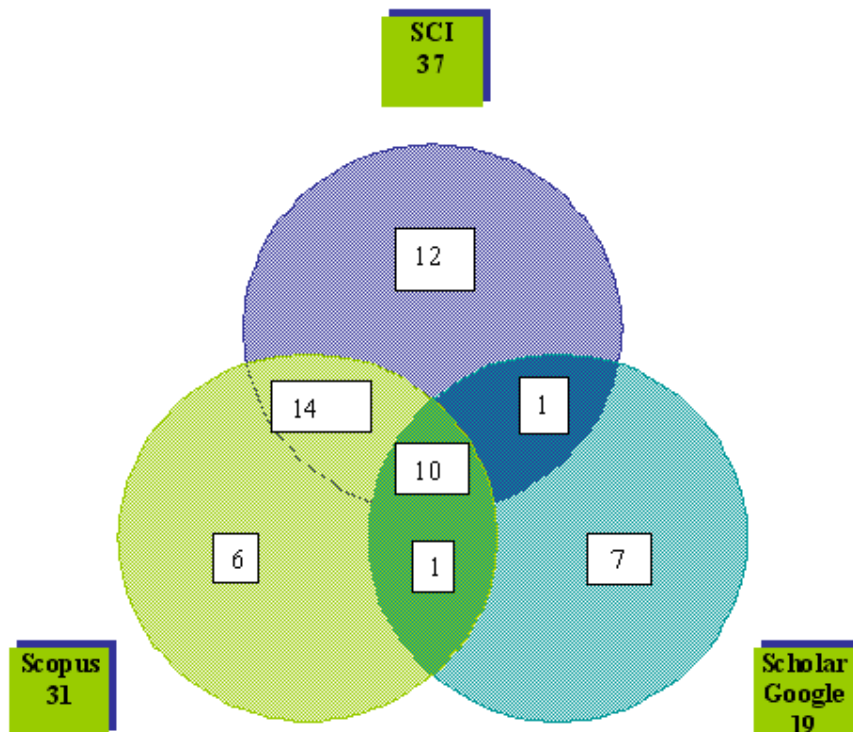
Tabela 3. Wyszukiwanie szczegółowe* dla jednego z autorów

Baza	GG		
	Liczba publ.	Liczba cytowań	h
SCI	37	233	11
Scopus	31	131	10
Scholar Google	19	59	6

*Wyszukiwanie szczegółowe – dodane wartości, wyszukane dodatkowo zadaniem pytaniem, w przypadku bazy SCI jest to pierwszy autor, w przypadku bazy Scopus – wyszukiwanie poprzez pole REFAUTH i zawężanie wyniku poprzez dodatkowych autorów, zakres tematyczny itp. (poprzez dostępne narzędzia)

Źródło: Opracowanie własne.

W przypadku autora GG różnice są znaczne. Po szczegółowym wyszukiwaniu w jego cytowaniach pojawiają się trzy dobrze cytowane prace, co automatycznie podnosi jego wskaźnik h z 8 do 11 w bazie SCI, gdzie wynik był najwyższy. Przy hipotetycznym założeniu, że przyznanie tytułu naukowego wymaga wskaźnika $h > 10$, od sposobu przeprowadzenia analizy cytowań mogłby więc zależeć jego awans.



Rys. 2. Prace autora GG we wszystkich bazach

We wszystkich trzech bazach powtarza się 10 pozycji tego autora, a unikalnych jest kolejno dla każdej z baz: SCI – 12, Scopus – 6, Google – 7. Oznacza to, że ten autor ma 37 (SCI) + 6 (Scopus) + 7 (Google), czyli 50 unikalnych pozycji, a tego nie pokazuje żadna baza.

Samo porównanie baz pod względem zawartości tytułów czasopism jest niekiedy złudne. W bazach SCI i Scopus wygląda to tak:

Tabela 4. Porównanie zawartości baz SCI i Scopus – zawartość tytułów czasopism

SCI	Scopus	Wspólne
6687	21684	6330

Źródło: Opracowanie własne.

Wydawać by się mogło, że skoro jedna z baz ma ponad dwukrotnie więcej tytułów, to cytowań również będzie znacznie więcej. Tymczasem w przypadku wybranych przeze mnie autorów tak nie jest (zob. tabela 2). W przypadku autora MK wyraźnie przeważa baza SCI, natomiast w przypadku autora DJ – baza Scopus. Pomimo że wykazuje mniej cytowanych publikacji, liczba cytowań jest znacznie wyższa. Wyższy jest też współczynnik h . Różnica między autorami polega na tym, że jeden z nich (MK) ma w swoim dorobku więcej publikacji książkowych, a tych baza Scopus w wyszukiwaniu prostym nie uwzględnia. Różnica ta zmniejszyłaby się w wyszukiwaniu szczegółowym. Gdyby natomiast analiza dotyczyła autorów publikujących w czasopiśmie polskich, wyniki byłyby zupełnie inne. Poniżej zamieszczam porównanie artykułów z „Przeglądu Geologicznego” w tych trzech bazach. „Przegląd Geologiczny” nie jest indeksowany w SCI i widać wyraźnie, że obszerniej cytuje baza Scopus [5].

Tabela 5. Artykuły z „Przeglądu Geologicznego” z lat 1996–2003 o największej liczbie cytowań odnotowanych w bazie Scopus™ do końca czerwca 2005 r. oraz porównanie z danymi z SCI Expanded i Google Scholar

Liczba cytowań sensu lato w: Scopus/SCI Ex/Google***	46/8/6	42/11/3	22/4/0	21/6/0	19/6/2	12/3/1	14/4/0	12/2/0	11/1/0	15/1/0
--	--------	---------	--------	--------	--------	--------	--------	--------	--------	--------

Źródło [5].

Przeprowadzone badania pokazują więc, że do analizy cytowań najlepiej wykorzystać wszystkie trzy bazy – bazy SCI i Scopus zamiennie, jeśli mamy dostęp do obu, a jeśli musimy zdecydować się na jedną z nich, to powinniśmy wiedzieć, w której z baz dziedziny wiedzy uprawiane na uczelni oraz związani z nią autorzy będą pełniej prezentowani. Jeśli publikują w czasopismach z „listy filadelfijskiej”, lepsza będzie baza SCI, jeśli zaś w ich dorobku przeważają publikacje w czasopismach spoza listy, pełniejszy wykaz będzie w bazie Scopus. Baza Google Scholar we wszystkich porównaniach wypada niekorzystnie. Na pewno nie zastąpi dwóch pozostałych, ale doskonale je uzupełni i wzbogaci analizę o dodatkowe, zupełnie niespodziewane cytowania.

Literatura

- [1] Bakkalbasi N, Bauer K., Glover J., Wang L.: *Three options for citation tracking: Google Scholar, Scopus and Web of Science* [dokument elektroniczny]. Tryb dostępu: <http://www.pubmedcentral.nih.gov/picrender.fcgi?artid=1533854&blobtype=pdf> [dostęp 10 września 2007].
- [2] Jacso P.: *As we may search – Comparison of major features of the Web of Science, Scopus, and Google Scholar citation-based and citation-enhanced databases*. “Current Science” 2005 Vol. 89 No. 9, p. 1537–1547.
- [3] Meho Lokman I., Yang Kiduk: *Impact of Data Sources on Citation Counts and Rankings of LIS Faculty: Web of Science vs. Scopus and Google Scholar* [dokument elektroniczny]. Tryb dostępu: <http://www.slis.indiana.edu/faculty/meho/meho-yang-03.pdf> [dostęp 10 września 2007].
- [4] *Publish or Perish* [dokument elektroniczny]. Tryb dostępu: <http://www.harzing.com> [dostęp 9 września 2007].
- [5] Racki G.: *Najbardziej znane publikacje Przeglądu Geologicznego z lat 1996–2003 (na podstawie bazy Scopus™)*. „Przegląd Geologiczny” 2005 vol. 53 nr 10/1, s. 839–843.